Two State of Art Image Segmentation Approaches for Outdoor Scenes

Mr. Jenopaul.P[1] and Anju.J.A[2]

[1] PSN College of Engineering and Technology, Tirunelveli, Tamil Nadu

Received: 7 December 2012 Accepted: 31 December 2012 Published: 15 January 2013

Abstract

The main research objective of this paper is to detecting object boundaries in outdoor scenes of images solely based on some general properties of the real world objects. Here, segmentation and recognition should not be separated and treated as an interleaving procedure. In this project, an adaptive global clustering technique is developed that can capture the non-accidental structural relationships among the constituent parts of the structured objects which usually consist of multiple constituent parts. The background objects such as sky, tree, ground etc. are also recognized based on the color and texture information. This process groups them together accordingly without depending on a priori knowledge of the specific objects. The proposed method outperformed two state-of-the-art image segmentation approaches on two challenging outdoor databases and on various outdoor natural scene environments, this improves the segmentation quality. By using this clustering technique is to overcome strong reflection and over segmentation. This proposed work shows better performance and improve background identification capability.

Index terms—

1 Introduction

mage segmentation is the process of partitioning a digital image into multiple segments. One of the fundamental problem in computer vision is considered as image segmentation. The primary goal of image segmentation is to simplify or change the representation of an image into something that is more meaningful and easier to analyse [1]. In general, the outdoor scenes can be categorized into two namely, unstructured objects (e.g., sky, roads, trees, grass, etc.) and structured objects (e.g., cars, buildings, people, etc.). The unstructured objects mainly consists of backgrounds of images and structured objects consists of foreground of images. The background objects usually have nearly homogenous surfaces and are distinct from the structured objects in images. So many appearance based methods are used to achieve high accuracy in recognizing these background object classes [2], [3], [4].

The challenge for outdoor segmentation comes from the structured objects that are often composed of multiple parts, with each part having distinct surface characteristics. Without certain knowledge about an object, it is difficult to group these parts together [5], [6]. The research objective of this paper is to explore detecting object boundaries in outdoor scene images only based on some general properties of the real-world objects, such as perceptual organization laws, without depending on a priori knowledge of the specific objects.

Perceptual organization plays an important role in human visual perception. Perceptual organization, refers to the basic capability of the human visual system to derive relevant groupings and structures from an image without prior knowledge of its contents. The Gestalt psychologists summarized some underlying principles (e.g., proximity, similarity, continuity, symmetry, etc) that lead to human perceptual grouping. The classic Gestalt laws pointed out that convexity also plays an important role on perceptual organization because many real-world objects such as buildings, vehicles, and furniture tend to have convex shapes. These can be summarized by a

single principle, i.e., the principle of nonacidentalness, which means that these structures are most likely produced by an object or process, and are unlikely to arise at random [7].

For applying Gestalt laws to real world applications there are several challenges. One of challenge is to find quantitative and objective measures of these grouping laws. The Gestalt laws are in descriptive forms. Therefore, one needs to quantify them for scientific use. Another challenge consists of finding a way to combine the various grouping factors since object parts can be attached in many different ways. Under different situations, different laws may be applied. Therefore, a perceptual organization system requires combining as many Gestalt laws as possible. The greater the number of Gestalt laws incorporated, the better chance the perceptual organization systems may apply appropriate Gestalt laws in practices. Ren [8] developed a probabilistic model of continuity and closure built on a scale-invariant geometric structure to estimate object boundaries. Jacobs emphasized that convexity plays an important role in perceptual organization and, in many cases, overrules other laws such as closure.

The main contribution of this paper is a developed perceptual organization model (POM) for boundary detection. The POM quantitatively incorporates a list of Gestalt laws and therefore is able to capture the nonaccidental structural relationships among the constituent parts of a structured object. With this model, we are able to detect the boundaries of various salient structured objects under different outdoor(D D D D D D D D)

Year environments.The proposed method outperformed two state-of-the-art studies [9], [10] on two challenging image databases consisting of a wide variety of outdoor scenes and object classes.

2 Methodology

The proposed system consists of three main steps for recognizing the common background and foreground objects.

3 a) Background Identification in Outdoor Natural Scenes

The objects seeming in natural scenes can be roughly divided into two categories namely, unstructured and structured objects. Unstructured objects typically have nearly similar surfaces, whereas structured objects typically consist of several essential parts, with each part having distinct appearances in their color, texture, etc. The common backgrounds in outdoor natural scenes are those unstructured objects such as skies, roads, trees, and grasses and these objects have low visual variability in most cases and are distinct from other structured objects in an image. For instance, a sky commonly has a identical form with blue or white colours; a tree or a grass usually has a textured presence with green colours. Hence, these background objects can be precisely predictable only based on appearance data. Assume if we use a bottom-up segmentation method to segment an outdoor image into uniform regions. Then, some of the regions must belong to the background objects. To recognize these background regions, we use a technique similar [2].

4 b) Perceptual Organization Model (POM)

Most images consist of background and foreground objects and these foreground objects are structured objects that are often composed of multiple parts, with each part having distinct surface characteristics. Assume that we can use a bottom-up method to segment an image into uniform patches, then most structured objects should be oversegmented to multiple parts. After the background patches are identified in the image, the majority of the remaining image patches correspond to the constituent parts of

5 READ IMAGE

6 PREPROCESSING FILTER BANK

7 ADAPTIVE GLOBAL CLUSTERING FOREGROUND EXTRACTION

8 ADABOOST

9 PERCEPTUAL ORGANIZATION MODEL

The key for this method is to use textons to represent object appearance information. The term texton is first presented for describing human textural perception. The whole textonization process proceeds as follows: First, the training images are converted to the perceptually uniform CIE color space. Then, the training images are convolved with a 17-D filter bank. We use the same filter bank as that in, which consists of Gaussians at scales 1, 2, and 4; the and derivatives of Gaussians at scales 2 and 4; and Laplacians of Gaussians at scales 1, 2, 4, and 8. The Gaussians are applied to all three color channels, whereas the other filters are applied only to the luminance channel. By doing so, we obtain a 17-D response for each training pixel. The 17-D response is then augmented with the CIE channels to form a 20-D vector. After augmenting the three color channels, we can achieve slightly higher classification accuracy [3]. Then, the Euclidean-distance -means clustering algorithm is performed on the 20-D vectors collected from the training images to generate cluster centers. These cluster

centers are called textons. Finally, each pixel in each image is assigned to the nearest cluster center, producing the texton map. After this textonization process, each image region of the training images is represented by a histogram of textons. We then use these training data to train a set of binary Adaboost classifiers to classify the unstructured objects (e.g., skies, roads, trees, grasses, etc.). to achieve high accuracy on classifying these background objects in outdoor images.

10 Global Journal of Computer Science and Technology

Volume XIII Issue II Version I8 (D D D D)

Year structured objects. The challenge here is how to piece the set of constituted parts of a structured object together to form a region that corresponds to the structured object without any object-specific knowledge of the object. To tackle this problem, we develop a POM. Accordingly, our image segmentation algorithm can be divided into the following three steps.

? Given an image, use a bottom-up method to segment it into uniform patches.

? Use background classifiers to identify background patches.

? Use POM to group the remaining patches (parts) to larger regions that correspond to structured objects or semantically meaningful parts of structured objects. We now go through the details of our POM. Even after background identification, there is still a large number of parts remaining. Different combinations of the parts form different regions. We want to use the Gestalt laws to guide us to find and group these kinds of regions. Our strategy is that, since there always exist some special structural relationships that obey the principle of nonaccidentalness among the constituent parts of a structured object, we may be able to piece the set of parts together by capturing these special structural relationships. The whole process works as follows: We first pick one part and then keep growing the region by trying to group its neighbors with the region. The process stops when none of the region's neighbors can be grouped with the region. To achieve this, we develop a measurement to measure how accurately a region is grouped. The region goodness directly depends on how well the structural relationships of parts contained in the region obey Gestalt laws. In other words, the region goodness is defined from perceptual organization perspective. With the region measurement, we can go find the best region that contains the initial part. In most cases, the best region corresponds to a single structured object or the semantically meaningful part of the structured object.

11 c) Image Segmentation Algorithm

The POM can capture the special structural relationships that obey the principle of nonaccidentalness among the constituent parts of a structured object. To apply the proposed POM to realworld natural scene images, we need to first segment an image into regions so that each region approximately corresponds to an object part. In this implementation, Felzenszwalb and Huttenlocher's approach [11] are used to generate initial superpixels for an outdoor scene image. We select this method because it is very efficient and the result of the method is comparable to the meanshift algorithm [12]. To further improve the segmentation quality, we apply a segment-merge method on the initial superpixels to merge the small size regions with their neighbors. These small size regions are often caused by the texture of surfaces or by the inhomogeneous portions of some part surfaces. Since these small size image regions contribute little to the structure information of object parts, we merge them together with their larger neighbors to improve the performance of our POM. In addition, if two adjacent regions have similar colors, we also merge them together. By doing so, we obtain a set of improved superpixels. Most of these improved superpixels approximately correspond to object parts. We now turn to the image segmentation algorithm.

Given an outdoor scene image, we first apply the segment-merge technique described above to generate a set of improved superpixels. Most of the superpixels approximately correspond to object parts in that scene. We build a graph to represent these superpixels: Let be an undirected graph. Each vertex corresponds to a superpixel, and each edge corresponds to a pair of neighboring vertices. We then use our background classifiers are divide into two parts: backgrounds such as sky, roads, grasses, and trees and structured parts. We then apply our perceptual organization algorithm at the beginning, all the components in are marked as unprocessed. Then, for each unprocessed component to detect the best region that contains vertex . Region may correspond to a single structured object or the semantically meaningful part of a structured object. We mark all the components comprising as processed. The algorithm gradually moves from the ground plane up to the sky until all the components in are processed. Then, we finish one round of perceptual organization procedure and use the grouped regions in this round as inputs for the next round of perceptual organization on. At the beginning of a new round of perceptual organization, we merge the adjacent components if they have similar colors and build a new graph for the new components. This perceptual organization procedure is repeated for multiple rounds until no components in can be grouped with other components. In practice, we find that the result of two rounds of grouping is good enough in most cases. At last, in a post process procedure, we merge all the adjacent sky and ground objects together to generate final segmentation. Year objects such as buildings, signs, cars, people, cows, and sheep. This data set provides ground truth object class segmentations that associate each region with one of eight semantic classes (sky, tree, road, grass, water, building, mountain, or foreground). In addition, the object class labels, the ground truth object segmentations that associate each segment with one physical object, are also provided. Following the same setup we randomly split the data set into 572 training images and 143 testing

images. Gould09 data set also used superpixels as a starting point. We used the normalized cut algorithm to generate 400 superpixels for use in the Gould09 method. The Gould09 method is a slight variant of the baseline method and achieved comparable result against the relative location prior method in Shotton's method and Yang's method on the MSRC-21 data set. Gould09 is trained on the training set and tested on the testing set. We first use the training images to train five background classifiers for background identification. Then, we test our POM method on both the testing set and the full GDS data set. We choose the method proposed by Martin as the measurement for segmentation accuracy.

12 III.

13 Experimental Results

The segmentation accuracy score is defined as where and represent the set of pixels in the ground truth segment of an object and the machine-generated object segment, respectively. Because all the images in this data set are downsized to 320 pixels 240 pixels, we set the parameters of Felzenszwalb's algorithm to small values to generate the initial superpixels from the input images. We found that Felzenszwalb's algorithm with this set of parameters works well for small size images . We set parameters for our POM and we used the 572 training images to learn five binary Adaboost classifiers to identify five background object classes (i.e., sky, road, grass, trees, and water). This compares the performance of our method. With that of the baseline method (Gould09) on the GDS. The segmentation accuracy measurement is based on the average value. For each class, the score is averaged over all the salient object segments in the class. For overall objects, the score is averaged over all the detected salient object segments. If the size of a ground truth object segment is smaller than 0.5% of the image size, it is not a salient object and will not be accounted for in the segmentation accuracy. In total, we detected 2757 salient objects from143 testing images and, on average, 19 objects per image. We are able to achieve an average improvement of 16.2% over the performance of the Gould09 method. Among 2757 salient objects detected in the testing images, the structured objects (buildings foregrounds) account for 52.6%. Our method significantly outperforms the Gould09 method on segmenting the structured objects. For the full data set, we detected 13 430 salient objects from 715 images and, on average, 18.8 objects per image. Structured objects account for 54.8% of the total detected salient objects.

For the structured objects, POM does not gain any prior knowledge from training images. Our POM achieves very stable performance on segmenting the difficultly structured objects on the full data set. This shows that our POM can successfully handle various structured objects appearing in outdoor scenes. Pixellevel accuracy reflects how accurate the classification is for multiclass segmentation methods. Pixel-level accuracy is computed as the percentage of image pixels correct class label. Our POM is not a multiclass segmentation method because it does not label each pixel of an image with one of eight semantic classes as Gould09. Therefore, our POM does not have pixel-level accuracy. Gould09 seems to be adaptable to the variation of the number of semantic classes. The method achieved 70.1% pixel-level accuracy on the 21class MSRC database according to and achieved impressive 75.4% pixel-level accuracy on the 8-class GDS. However, the foreground class in GDS includes a wide variety of structured object classes such as cars, buses, people, signs, sheep, cows, bicycles, and motorcycles, which have totally different appearance and shape characteristics. This makes training an accurate classifier for classifying the foreground classes difficult. As a result, the Gould09 method cannot handle complicated environments where multiple foreground objects may appear close to each other. In such cases, the Gould09 method often labeled the whole group of physically different object instances such as people, car, and sign as one continuous foreground class region. This affects the performance of Gould09 on the objectlevel segmentation. If the foreground class can be further divided into more semantic object classes, the performance of the Gould09 method can be expected to improve on the GDS. The small number of semantic classes does not affect our method. Our method only requires identifying five background object classes (i.e., sky, trees, road, grass, and water). The remaining object classes are treated as structured objects. b) Berkeley Segmentation Data Set POM image segmentation method can be evaluated by using Berkeley segmentation data set (BSDS). BSDS contains a training set of 200 images and a test set of 100 images. For each image, BSDS provides a collection of hand-labeled segmentations from multiple human subjects as ground truth. BSDS has been widely used as a benchmark for many boundary detection and segmentation algorithms in technical literature. We directly evaluate our POM method on the test set of BSDS. The sizes of images in this data set are 481 321, which are larger than the sizes of images in GDS. We use larger parameters for Felzenszwalb's algorithm to generate the initial superpixels for an input image. We use the same background classifiers trained in the GDS data set to identify background objects in this data set. Examples of our POM segmentation algorithm on the BSDS data set.

The region-based segmentation accuracy measurement is still. For each image, BSDS provides a collection of multiple human-labeled segmentations. For simplicity, we only select the first human-labeled segmentation of the collection as ground truth for the image. The score is averaged over all the salient object segments. If the size of a ground truth segment size is smaller than % 0.5 of the image size, it is not a salient object and will not be accounted for segmentation accuracy. In total, we detect 681 salient objects from 100 images and, on average, 6.8 objects per image. Our POM achieved an averaged segmentation accuracy score of 53% on the test set of BSDS. For the boundarybased measurement, we use the precision-recall framework recommended by BSDS. A precision-recall curve is a parameterized curve that captures the trade off between accuracy and noise. Precision

is the fraction of detections that are true boundaries, whereas recall is the fraction of true boundaries that are detected. Thus, precision is the probability that the segmentation algorithm's signal is valid, and recall is the probability that the ground truth data is detected. These two quantities can be combined in a single quality measure, i.e., F-measure, defined as the weighted harmonic mean of precision and recall. Boundary detection algorithms usually generate a soft boundary map for an image.

IV.

14 Conclusion

The main contribution of this paper is to develop a perceptual organization model for extracting background and foreground images of an object. Our experimental results show that our future method outpaced two competing state-of-the-art image segmentation approaches and achieved good segmentation quality on two challenging outdoor scene image data sets. It is well accepted that segmentation and recognition should not be separated and should be treated as an interleaving procedure. In this method mainly follows the scheme and requires identifying some background objects as a starting point and compared to the large number of structured object classes. There are only a few common background objects in outdoor scenes and these objects have low visual variety and hence can be reliably recognized. After background objects are identified, we roughly know where the structured objects are and delimit perceptual organization in certain areas of an image. Our method can piece the whole object or the main portions of the objects together without requiring recognition of the individual object parts. In other words, for these object classes, our method provides a way to separate Year segmentation and recognition. This is the major difference between our method and other class segmentation methods that require recognizing an object in order to segment it. This paper shows that, for many fairly articulated objects, recognition may not be a requirement for segmentation. The geometric relationships of the constituent parts of the objects provide useful cues indicating the memberships of these parts. [1] [2] [3]

Figure 1: Figure 1 :

[1] F © 2013 Global Journals Inc. (US)
[2] F © 2013 Global Journals Inc. (US)
[3] F © 2013 Global Journals Inc. (US)

Figure 2:

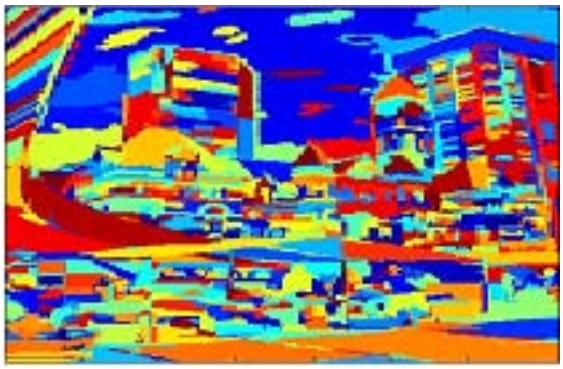
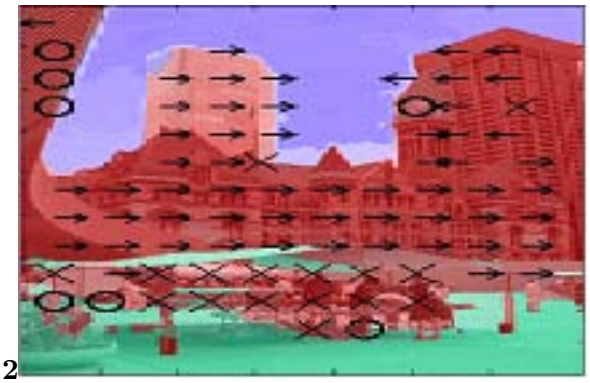


Figure 3: F



2

Figure 4: Figure 2 :

[IEEEWorkshop Perceptual Org. Comput. Vis., CVPR ()] , *IEEEWorkshop Perceptual Org. Comput. Vis., CVPR* 2004. p. .

[Int. J. Comput. Vis (2008)] , *Int. J. Comput. Vis* Oct. 2008. 80 (1) p. .

[Winn et al. ()] 'Categorization by learned universal visual dictionary'. J Winn , A Criminisi , T Minka . *Proc. IEEE ICCV*, (IEEE ICCV) 2005. 2 p. .

[Borenstein and Sharon] *Combining top-down and bottom-up segmentation*, E Borenstein , E Sharon . (in Proc)

[Gould et al. ()] 'Decomposing a scene into geometric and semantically consistent regions'. S Gould , R Fulton , D Koller . *Proc. IEEE ICCV*, (IEEE ICCV) 2009. p. .

[Felzenszwalb and Huttenlocher (2004)] 'Efficient graph-based image segmentation'. P Felzenszwalb , D Huttenlocher . *Int. J. Comput. Vis* Sep. 2004. 59 (2) p. .

[Rutishauser and Walther ()] 'Is bottom-up attention useful for object recognition?'. U Rutishauser , D Walther . *Proc. IEEE CVPR*, (IEEE CVPR) 2004. 2 p. .

[Ren et al. (2008)] 'Learning probabilistic models for contour completion in natural images'. X F Ren , C C Fowlkes , J Malik . *Int. J. Comput. Vis* May 2008. 77 (1-3) p. .

[Comaniciu and Meer (2002)] 'Mean shift: A robust approach toward feature space analysis'. D Comaniciu , P Meer . *IEEE Trans. Pattern Anal. Mach. Intell* May 2002. 24 (5) p. .

[Gould et al. (2008)] 'Multi-class segmentation with relative location prior'. S Gould , J Rodgers , D Cohen , G Elidan , D Koller . *Int. J. Comput. Vis* Dec. 2008. 80 (3) p. .

[Yang et al. ()] 'Multiple class segmentation using a unified framework over manshift patches'. L Yang , P Meer , D J Foran . *Proc. IEEE CVPR*, (IEEE CVPR) 2007. p. .

[Pantofaru et al. ()] 'Object recognition by integrating multiple image segmentations'. C Pantofaru , C Schmid , M Hebert . *Proc. ECCV*, (ECCV) 2008. p. .

[Shah] *Performance modeling and algorithm characterization for robust image segmentation*, S K Shah .

[Cheng et al. ()] 'Scene image segmentation based on perception organization'. C Cheng , A Koschan , D L Page , M A Abidi . *Proc. IEEE ICIP*, (IEEE ICIP) 2009. p. .

[Micusik and Kosecka ()] 'Semantic segmentation of street scenes by superpixel co-occurrence and 3-D geometry'. B Micusik , J Kosecka . *Proc. IEEE Workshop VOEC*, (IEEE Workshop VOEC) 2009.

[Shotton et al. ()] 'Semantic texton forests for image categorization and segmentation'. J Shotton , M Johnson , R Cipolla . *Proc. IEEE CVPR*, (IEEE CVPR) 2008. p. .

[Shotton et al. (2009)] 'Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context'. J Shotton , J Winn , C Rother , A Criminisi . *Int. J. Comput. Vis* Jan. 2009. 81 (1) p. .

[Gould et al.] *The STAIRVision Library*, S Gould , O Russakovsky , I Goodfellow , P Baumstarck , A Y Ng , D Koller . `http://ai.stanford.edu/~sgould/svl` (v2.3) 2009 [Online)

[Maire et al. ()] 'Using contours to detect and localize junctions in natural images'. M Maire , P Arbelaez , C C Fowlkes , J Malik . *Proc. IEEE CVPR*, (IEEE CVPR) 2008. p. .

[Jacobs (2003)] 'What makes viewpoint-invariant properties perceptually salient?'. D W Jacobs . *J. Opt. Soc. Amer. A, Opt. Image Sci* Jul. 2003. 20 (7) p. .